NOMA-based Cooperative Opportunistic Multicast Transmission Scheme for Two Multicast Groups: Relay selection and Performance Analysis

Yufang Zhang, Xiaoxiang Wang, Dongyu Wang, Yibo Zhang, Qiang Zhao, and Qian Deng

Abstract—The orthogonal frequency division multiple access (OFDMA)-based cooperative multicast (CM) technology realizing the intra-cooperation of multicast group (MG) has improved the performance in throughput and coverage. In the MG scenario, users are uniformly distributed. Due to the small effective transmission range (ETR) of relay user (RU), the cooperation is hindered, which limits the performance of CM. Therefore, non-orthogonal multiple access (NOMA)-based two MGs joint cooperative opportunistic multicast scheme (termed as COM-NOMA) is proposed. In COM-NOMA scheme, two MGs are served as one quasi-MG. Firstly, on the same frequency band, base station (BS) opportunistically multicasts the superposition signal, which is formed by two MGs’ signals, to the users with good channel conditions in the two MGs using NOMA way. Secondly, some successfully receiving users are selected as RUs. The superposition signal is forwarded to the other users by Device-to-Device multicast (D2MD) way. Then, according to the coverage ratio of BS and the decode condition of NOMA, the signal to interference and noise ratio (SINR) threshold for RU selection is given. Finally, the spectral efficiency of COM-NOMA is analyzed and the approximate expression of system coverage ratio is derived. The simulation results show that, spectral efficiency at unit transmit power (spectral efficiency per watt) and system coverage ratio is improved in comparison with other OFDMA-based CM schemes.

Index Terms—Cooperative multicast, coverage ratio, non-orthogonal multiple access, relay selection, signal to interference and noise ratio threshold.

I. INTRODUCTION

Due to the development of multi-media services, mobile data increases dramatically. Providing subscribers with high-quality multimedia service on limited bandwidth becomes a hot topic and has been paid much attention recently in both the industry and the academic. For the multimedia video-centric applications [1] (e.g., video on demand, video games, live video streaming, video conference, video surveillance, etc.) that have high popularity, users request numerously during popular time. To maximize system throughput, operators give preference to multicast [2]. It can be seen that multicast services are still the leading drive in mobile data increase [3]. It means that a large quantity of spectrum resources are occupied by multicast services. The pressure on spectrum resources caused by multicast will become a big issue in future.

Orthogonal frequency division multiple access (OFDMA)-based two-stage cooperative multicast (CM) as an advanced multicast scheme has already been researched in improving system throughput and enhancing coverage1. In one time slot, the process of multicast is divided into two stages. In the first stage, through opportunistic multicast technology, base station (BS) multicasts data at high rate so that users with good channel conditions successfully receive (termed as SUs). The other users who are failed to receive are unsuccessful users (termed as USUs). In the second stage, some SUs are selected as relay users (RUs) to serve USUs.

In [4], a certain proportion (e.g., 50%) of users in a multicast group (MG) successfully receive information from BS in the first stage, and then each USU receives signals forwarded by all SUs simultaneously. To improve energy efficiency, when USU is in the overlapping coverage area of multiple SUs, it only receives information from these SUs [5]. In [6], RU is placed on the fixed position in the cell. In order to minimize system average outage probability, the optimal location is found to place RUs in the proposed genie-aided cooperative multicast scheme [6]. A try-best RU selection scheme in [8] chooses the closest SU to USU as relay. Those cooperative multicast schemes implement intra-group collaboration of MG. However, the transmission power of RU is low, and the radius of its effective transmission range (ETR) is far less than the radius of the cell. When MG users are uniformly distributed, SUs sparsely locate in the cell. It is difficult for RU to cover USUs. In the first stage, the BS performs opportunistic multicast through modulation and coding scheme (MCS) selection according to the minimal channel gain of the users with good channel conditions (i.e., SUs). Therefore, after forwarding in the second stage, when the USU’s reception SINR is greater than the equivalent signal to interference and noise ratio (SINR) of the first-stage opportunistic multicast, demodulation can be successfully performed. However, in the above studies, SINR threshold for RU selection is equal to average value of the first-stage multicast reception SINR in every time slot, which is set in MCS. It doesn’t make sure that USU successfully receives in the instantaneous channel

1The coverage ratio is the ratio of the users who have successfully received among the total MG users
Non-orthogonal multiple access (NOMA) as a more efficient multiple access technology in spectral efficiency improvement and user fairness compared with OFDMA, has attracted a lot of attention [9]. By utilizing the power domain rather than the conventional time and frequency domains, NOMA can significantly improve network throughput [10]. The essential idea of NOMA is that multiple users can share the same frequency resources and use different power levels [11].

NOMA-based cooperative relaying is introduced into multicast in a cognitive network. It utilizes the characteristic of NOMA that several signals share the same frequency band to enhance the capacity and fairness. Every receiver can receive all signals from the same frequency band. One unicast user (i.e., the primary user) locates closely to the edge of the cell and its channel condition is poor, while the multicast users (i.e., the secondary users) seat closely to the BS such that they have good channel conditions. In the first stage, unicast user and some multicast users can access the same licensed spectrum through NOMA technology due to the difference in channel conditions between two kinds of users. In the second stage, the multicast users selected as relays forward data to the unicast user [12]–[15]. Even though MG users have already subscribed and paid for the service, as the secondary users, their performance cannot be guaranteed.

All the aforementioned research efforts on multicast are limited into a single MG. A joint transmission scheme based on NOMA to further increase the spectral efficiency, which aims at two MGs, is proposed in [16]. But the scheme serves all MG users. On one hand, its system throughput is still limited by the user with the worst channel condition, which is also the bottleneck problem of multicast and has not been settled. On the other hand, via successive interference cancellation (SIC), the other MG’s signal superimposed on the same frequency band with the one MG’s signal is cancelled directly. The superposition signal in NOMA can not be fully utilized.

Here, we introduce the NOMA technology for two MGs into the CM. A two-stage COM-NOMA is researched, since its lower signaling overhead and complexity than the multi-stage one. In the first stage, BS sends superposition signal to the users of two MGs who own good channel conditions. In the second stage, the SUs that are selected as RUs directly forward to the remaining users through D2MD. RUs transmit data simultaneously, same as in [6], [17], [18]. Although the transmission power of the RU is low, the small ETR means small path loss. High reception SINRs can be achieved by USUs. The main contributions of this paper are threefold:

- The NOMA-based cooperative opportunistic multicast scheme serving two MGs on the same frequency band is given. In the first stage, the signals of two MGs are superimposed on the power domain. BS accesses the users with good channel conditions in the two MGs. In the second stage, some SUs are selected to forward the superposition signal for USUs via D2MD technology. Simulation results show that the proposed scheme possesses high power efficiency and enhances coverage ratio with raising number of MG users.

- For the instantaneous channel conditions, according to the number of users in each MG, two methods of achieving SINR threshold for RU selection are given. When the number is small, COM-NOMA scheme makes the first-stage real-time multicast SINR as threshold for USUs to select RUs to ensure they can successfully decode in the second stage. When the number is large, in order to reduce the frequency that BS computes and broadcasts the thresholds, the closed-from expression of SINR threshold is derived, according to the extreme value theory and statistical analysis. When the first-stage coverage ratio is known, the corresponding threshold for relay selection can be calculated. Simulation results prove the calculated threshold can satisfy USU’s decode condition.

- The approximate closed-form expression of the COM-NOMA coverage ratio is derived. In the process of the second-stage coverage ratio calculation, in order to reduce the complexity of integral calculation, the circular area covered by the RU is approximated to its circumscribed sector area. Simulation results show that the approximate coverage ratio and the simulated one match well. Furthermore, the power allocation factor (PAF), as a key parameter of NOMA, scarcely impacts on the coverage ratio of COM-NOMA scheme.

The rest of this paper is organized as follows. We describe the two-stage cooperative opportunistic multicast based on NOMA in Sec. II. In Sec. III, the second-stage SINR threshold is investigated for USU to select RU. In Sec. IV, we analyze the system coverage ratio and give its approximate close-form expression. The complexity of COM-NOMA scheme is also analyzed in this section. Numerical simulation results and analysis are given in Sec. V. Finally, we conclude this paper and highlight our findings in Sec. VI.

II. SYSTEM MODEL OF COM-NOMA AND SPECTRUM EFFICIENCY ANALYSIS

Consider a downlink two-stage COM-NOMA framework consisting of one BS, two MGs as illustrated in Fig. 1. The two MGs are denoted by $MG_i$, $i \in \{1, 2\}$. The number of users in $MG_i$ is $M_i$. They are uniformly distributed with a probability density function of distance $\gamma_i$ from the BS, which is expressed as $f_\gamma(\gamma_i) = \frac{2\gamma_i}{R^2}$, $0 < \gamma_i \leq R$, where $R$ is the radius of the cell. They are equipped with one antenna. BS gets all channel state information (CSI). It is assumed that CSI will not change during a transmission.
time slot $T$, but be different in different time slots. The channel gains of users in $MG_i$ are in descending order denoted by set $H_i = \{H_{i,1}, H_{i,2}, \cdots, H_{i,k_i}, \cdots, H_{i,M_i}\}$, $k_i \in [1, M_i]$, where $H_{i,k_i} = \|h_{i,k_i}\|^2 \gamma_{i,k_i}$, $h_{i,k_i}$ is the complex channel coefficient for the link from BS to user $k_i$ and follows circularly symmetric complex Gaussian distribution $CN \sim (0, 1)$. $\gamma_{i,k_i}$ denotes the Euclidean distance between them. $\beta$ is the path loss parameter. Time slot is divided into two stages, i.e., $T_1$ and $T_2$, $T_1 + T_2 = T$. The work in [19] has proved that the maximal system throughput can be achieved, when $T_1 = T_2 = T/2$. It is also assumed in this paper. Downlinks from BS to every users and between any two users are subject to Rayleigh fading, propagation path loss and additive white Gaussian noise (AWGN). COM-NOMA still occupies the bandwidth originally allocated to the two MGs in the OFDMA system.

In the first stage, through opportunistic multicast BS only serves a fraction of MG users who have good channel conditions. In the transmission of BS, two MGs’ signals are transmitted simultaneously by NOMA signaling on the same frequency band. The remaining users with bad channel conditions give up receiving the signal that leads to failed decoding. In the second stage, BS mutes and some of the SUs are selected as RUs, as illustrated in Fig.1. It is assumed that all users are accommodating. They are willing to forward data for other users.

**A. The First Stage Transmission**

In the first stage of COM-NOMA, BS opportunistically multicasts to the users with good channel conditions of two MGs at coverage ratio $C_1$ using the joint transmission scheme proposed in [16]. They are denoted by set $S_i$, when they belong to $MG_i$. $S_i = \{SU_{i,1}, SU_{i,2}, \cdots, SU_{i,m_i}, \cdots, SU_{i,s_i}\}$, $s_i = C_1M_i, m_i \in [1, s_i]$. The corresponding channel gain set is $H_{S_i} = \{H_{i,1}, H_{i,2}, \cdots, H_{i,m_i}, \cdots, H_{i,s_i}\}$. Without loss of generality, $H_{1,s_1} > H_{2,s_2}$ is assumed here. $H_{S_i}$ is divided into three subgroups, $H_{S_{i,b}}$, $H_{S_{i,m}}$ and $H_{S_{i,w}}$, by Maximum and Minimum group method (MMGM) as is shown in **Algorithm 1**. And $H_{S_{i,b}} > H_{S_{i,m}} > H_{S_{i,w}}$. In Fig.2, the case of $H_{1,s_1} > H_{2,s_2}$ and $H_{1,1} > H_{2,1}$ is shown. The corresponding user subsets are $S_{i,b}^1$, $S_{i,m}^1$ and $S_{i,w}^1$

Because $H_{S_{i,w}} < H_{S_{i,b}}$, BS transmits the superposition signal in NOMA to their corresponding user subgroup pair (SP), $S_{i,b}^w$ and $S_1 (S_1 = S_{i,m}^1 \cup S_{i,w}^1)$. And the transmission rate is decided by the least channel gain, $H_{2,s_2}$. The transmission signal is $X_T = \sqrt{\alpha_1P}x_1 + \sqrt{\alpha_2P}x_2$, where $P$ is the transmission power of BS. $\alpha_1$ and $\alpha_2$ are the PAFs of
Algorithm 1 Maximum and Minimum group method

Require: Two MPs’ channel gains sets: \( \mathcal{H}_{S_i}, i = \{1, 2\} \)

Ensure: \( \mathcal{H}_{S_{b}}, \mathcal{H}_{S_{m}}, \mathcal{H}_{S_{w}} = \mathcal{H}_{S_i} \)

1: if \( \exists H_{1,m_1} < H_{2,s_2} \) then
2: \( \mathcal{H}_{S_{1,w}} = \{ H_{1,m_1} | H_{1,m_1} < H_{2,s_2} \} \)
3: else if \( \exists H_{1,m_1} = H_{2,s_2} \) then
4: \( \mathcal{H}_{S_{1,w}} = \{ H_{1,m_1} | H_{1,m_1} = H_{2,s_2} \} \)
5: else
6: \( \mathcal{H}_{S_{2,w}} = \{ H_{2,m_2} | H_{2,m_2} < H_{1,s_1} \} \)
7: \( \mathcal{H}_{S_{1,w}} = \emptyset \)
8: end if
9: if \( \exists H_{1,m_1} > H_{2,1} \) then
10: \( \mathcal{H}_{S_{1,b}} = \{ H_{1,m_1} | H_{1,m_1} > H_{2,2} \} \)
11: \( \mathcal{H}_{S_{2,b}} = \emptyset \)
12: else if \( \exists H_{1,m_1} = H_{2,2} \) then
13: \( \mathcal{H}_{S_{1,b}} = \{ H_{1,m_1} | H_{1,m_1} = H_{2,2} \} \)
14: \( \mathcal{H}_{S_{2,b}} = \{ H_{2,1} \} \)
15: else
16: \( \mathcal{H}_{S_{2,b}} = \{ H_{2,m_2} | H_{2,m_2} > H_{1,1} \} \)
17: \( \mathcal{H}_{S_{1,b}} = \emptyset \)
18: end if

Given \( H_{1,s_1} > H_{2,s_2}, SINR_{S_1}^{S_2} > SINR_{S_2} \). So they can decode the interference signal \( x_2 \) firstly. After SIC, \( x_2 \) is cancelled and \( x_1 \) is obtained. Thanks to their larger SINRs than \( SINR_{S_2, S_1} \), SUs in \( S_1 \) decode their demand signal \( x_1 \) successfully. The equivalent reception SINR of \( x_1 \) after SIC is

\[
SINR_{S_1} = \frac{\alpha_1 P|h_{1,m_1}|^2 \gamma_{1,s_1}^{-\beta}}{2N_0}.
\] (5)

Following the analysis in [16], the throughput of the opportunistic multicast in the first stage is

\[
R_1 = 2B \sum_i C_i M_i \log_2 (1 + SINR_{S_i}).
\] (6)

In summary, \( MG_1 \) and \( MG_2 \) are regarded as one quasi-MG in the first stage. BS transmits the signal \( X_T \) to the quasi-MG by opportunistic multicast with fixed coverage ratio \( C_1 \). After the first stage, \( SU_{i,m} \) out of \( S_i \) who has already simultaneously received \( x_1 \) and \( x_2 \), can be enlisted as candidate relays for USUs.

B. The Second Stage Transmission

In the second stage, the selected RU forwards the first-stage reception signals in Device-to-Device multicast (D2MD) [20] to USU. The set of remaining USUs is denoted by \( U \).

One important fact in D2D communication is that the radius of ETR (e.g., 100m) [21], [22] is much smaller than that of cell due to the low transmission power of the RU. Therefore, only SUs in the vicinity of a USU (that is, SUs in the USU’s ETR) can serve as candidate relays. However, the small radius of ETR also means USU has probability to get higher SINR than first-stage SUs so that the demand signal can be decoded from the superposition signal. The radius of ETR is denoted by \( R_0 \), \( R_0 < < 0 \).

From the analysis of previous subsection, the larger received SINR than \( SINR_{S_2} \) guarantees successfully decoding. Therefore, the SINR threshold \( \sigma_0 \) for RU selection is not smaller than the minimum SINR in the first-stage opportunistic multicast, i.e., \( \sigma_0 \geq SINR_{S_2} \). The RU set of \( u \) is denoted by \( R_u = \{ SU_{i,m} | \gamma_{m_i,u} \geq R_0, SINR_{m_i,u}^{S_2} \geq \sigma_0 \} \). \( \gamma_{m_i,u} \) is the distance from \( u \) to its RU \( SU_{i,m} \). \( SINR_{m_i,u}^{S_2} \) denotes the reception SINR of \( u \) from \( SU_{i,m} \), for signal \( x_2 \).

The reception signal of \( u \) from \( SU_{i,m} \) is

\[
Y_u = h_{m_i,u}(\sqrt{\alpha_1 P_D} x_1 + \sqrt{\alpha_2 P_D} x_2) + \omega_n.
\] (7)

\( P_D \) is RU’s transmission power, which is much smaller than \( P \). The reception SINR is

\[
SINR_{m_i,u}^{S_2} = \frac{\alpha_2 P_D|h_{m_i,u}|^2 \gamma_{m_i,u}^{-\beta}}{\alpha_1 P_D|h_{m_i,u}|^2 \gamma_{m_i,u}^{-\beta} + 2N_0}.
\] (8)

The interference signal has already been cancelled. SINR here is to simplify writing of (5).
The RU selection and transmission strategy is shown in Algorithm 2. The USUs served by the same relay SU$_{i,m_i}$ consist a D2MD group, which is denoted by D$_{M_{m_i}}$. The number of USUs in D$_{M_{m_i}}$ is |D$_{M_{m_i}}$|. The set of the D2MD transmitters is denoted by R$_{D2MD}$. R$_{D2MD}$ is a union of all R$_{u}$.

The RU selection and transmission strategy is shown in Algorithm 2.

Algorithm 2 RU Selection and Transmission Strategy

1: BS broadcasts a training signal containing a quantized version of σ$_0$ to the users in the two MGs.
2: Every MG user broadcasts a flag message, Success or Failure according to it is a SU or not. If the flag message is Failure, the SU$_{i,m_i}$, measures the channel gain between the SU $u$ and itself, which is denoted as $h_{m_i,u}$. As thus, the CSI of other MG users within the ETR are obtained.
3: SU$_{i,m_i}$ detects obtained channel gain of every USU. If there are channel gains of USUs making the value of (8) larger than σ$_0$. This SU$_{i,m_i}$ is selected as a RU. Those USUs who make the value of (8) larger than σ$_0$ are the D2MD members served by SU$_{i,m_i}$.
4: Every selected RU forwards reception signal in D$_{2MD}$.

Every selected RU forwards reception signal in D$_{2MD}$.

The USUs served by the same relay SU$_{i,m_i}$ consist a D2MD group, which is denoted by D$_{M_{m_i}}$. The number of USUs in D$_{M_{m_i}}$ is |D$_{M_{m_i}}$|. The set of the D2MD transmitters is denoted by R$_{D2MD}$. R$_{D2MD}$ is a union of all R$_{u}$.

The RU selection and transmission strategy is shown in Algorithm 2.

For signal $x_2$, the equivalent reception SINR in D$_{M_{m_i}}$ is

$$SINR_{D_{M_{m_i}}} = \min_{u \in D_{M_{m_i}}} SINR_{m_i,u}^{x_2}. \tag{9}$$

U$_{i,m_i}$ is the collection of USUs out of D$_{M_{m_i}}$. Its subscript $i$ represents the MG that the USU belongs to. Its superscript $m_i$ represents the number of USUs in U$_{i,m_i}$. |U$_{i,m_i}$| represents the number of USUs in U$_{i,m_i}$. $u_{j,m_i}$ is the USU in U$_{i,m_i}$. The D2MD reception SINRs of $u_{j,m_i}$ are

$$SINR_{1}^{i,m_i} = \min_{u \in U_{i,m_i}} \left(\frac{\alpha_1 P_d h_{m_i,u}^2}{2 N_0}\right)^{2-\beta}.$$ \tag{10}

$$SINR_{2}^{m_i} = SINR_{D_{M_{m_i}}}. \tag{11}$$

It is assumed that the cyclic prefix (CP) is longer than the maximum delay of the equivalent multipath channel, all the received signals arriving within the CP duration can be added up to construct a stronger signal, which is denoted as CP combining [6]. So the throughput of the second stage is

$$R_2 = 2B \sum_{SU_{i,m_i} \in R_{D2MD}} \sum_i |U_{i,m_i}| \log_2 (1 + SINR_{i,m_i}). \tag{12}$$

The spectral efficiency of COM-NOMA is

$$E = \frac{1}{2} \frac{(R_1 + R_2)}{2B}. \tag{13}$$

III. Analysis of SINR Threshold for RU Selection

In this section, we analyze the SINR threshold σ$_0$. In the second stage, if SU$_{i,m_i}$ is a RU of USU $u$, the distance between them should be less than $R_0$, and he (she) can provide USU $u$ with greater SINR than SINR$_{S_2}$, which is decided by the opportunistic multicast in the first stage. From the analysis in Sec. II-A, based on the instantaneous channel conditions, when the number of MG users is small, BS is able to compute and broadcast the threshold frequently, and σ$_0$ = SINR$_{S_2}$. However, when the number of MG users is large, it is a heavy load for BS. To reduce the burden of the BS and get effective SINR threshold, statistical property of SINR$_{S_2}$ is taken into account.

According to the extreme value theory and lemma [23], given coverage ratio $C_1$, the transmission rate per unit bandwidth of opportunistic multicast $R_{2,s_2}$, which is decided by the SU with the smallest channel gain in the first stage, follows normal distribution.

$$R_{2,s_2} \sim C \mathcal{N} \left(F_{R_{2,s_2}}^{-1} \left(1 - C_1 \right), \frac{\delta^2(C_1)}{N} \right).$$

So in COM-NOMA, let $Y = \frac{\gamma_{2,s_2}}{\gamma_{2,s_2} - \beta + 2}$. The CDF of X is

$$F_{X|\gamma_{2,s_2}}(x) = 1 - \exp(\lambda_{S_2} x). \tag{15}$$

So in COM-NOMA, let $y = \frac{\alpha_2 x + \beta}{\alpha_1 x + \beta}$. SINR$_{S_2} = y$. The conditional distribution of $Y = \frac{\alpha_2 x + \beta}{\alpha_1 x + \beta}$ is

$$F_{Y|\gamma_{2,s_2}}(y) = P(Y \leq y)$$

$$= P \left( X \leq \frac{2Y}{(\alpha_2 - \alpha_1)Y} \right)$$

$$= 1 - \exp \left( -\lambda_{S_2} \frac{2y}{(\alpha_2 - \alpha_1)Y} \right).$$ \tag{16}
and the marginal CDF of $Y$ is

$$F_Y(y) = \int_0^y f_{\gamma_{2,s_2}}(y) f_\gamma(\gamma_{2,s_2}) d\gamma_{2,s_2}$$

From (18) and (19), given a first-stage coverage ratio $C_1$, the mean value and variance of minimal reception rate per unit bandwidth $R_{2,s_2}$, that supports SUs to decode their demand signals can be obtained. According to thrice standard error principle of normal distribution, the rate threshold for relay selection in large number of MG users is set to

$$R_{\sigma_0}(C_1) = F_{R_{2,s_2}}^{-1}(1 - C_1 + 3 \cdot \delta(C_1) \sqrt{N}).$$

For the instantaneous channel conditions, the event that the first-stage multicast rate $R_{2,s_2}$ is greater than $R_{\sigma}$, can be considered as a small probability event. When the rate threshold is set to $R_{\sigma_0}$, it can be ensured that the reception rate of USU in D2MD group is greater than that of the first stage, so that the USU is able to decode the demand data successfully from the NOMA signaling.

According to Shannon theory, $y = 2^{r_{2,s_2}} - 1$. The CDF of $R_{2,s_2}$ is

$$F_{R_{2,s_2}}(r_{2,s_2}) = \int_0^{2^{r_{2,s_2}} - 1} F_Y(y) dy$$

where $\nu = \frac{2(2^{r_{2,s_2}} - 1)}{\alpha_2 - \alpha_1(2^{r_{2,s_2}} - 1)}$. So the PDF of $R_{2,s_2}$ is

$$f_{R_{2,s_2}}(r_{2,s_2}) = \frac{dF_{R_{2,s_2}}(r_{2,s_2})}{dr_{2,s_2}}$$

The first order Taylor series approximation of $\exp(-t)$ is difficult to be derived, to obtain further insights of $R_{\sigma_0}(C_1)$, its approximate expression is given, when $\frac{P}{N_0} \to \infty$.

$$F^A_{R_{2,s_2}}(r_{2,s_2}) = \int_0^r \left[ 1 - \exp\left(-\frac{N_0(\gamma_{2,s_2})}{P} \frac{2(2^{r_{2,s_2}} - 1)}{\alpha_2 - \alpha_1(2^{r_{2,s_2}} - 1)} \right) \right] \cdot \frac{2\gamma_{2,s_2}}{R^2} d\gamma_{2,s_2} \approx \int_0^{N_0} \frac{2(2^{r_{2,s_2}} - 1)}{P} \frac{2(2^{r_{2,s_2}} - 1)\gamma_{2,s_2}^{\frac{\beta}{\beta}+1} - \alpha_2 - \alpha_1(2^{r_{2,s_2}} - 1)}{R^2} d\gamma_{2,s_2}$$

For $t$ close to 0 is used here. The approximate expression of $f_{R_{2,s_2}}(r_{2,s_2})$ is

$$f^A_{R_{2,s_2}}(r_{2,s_2}) = \frac{dF^A_{R_{2,s_2}}(r_{2,s_2})}{dr_{2,s_2}} = \frac{N_0}{P} \frac{R^3}{\beta + 1} \cdot 4 \nu_1.$$ (23)

And

$$\nu = \frac{N_0(\gamma_{2,s_2})}{P} \nu_1 \exp(-\frac{N_0(\gamma_{2,s_2})}{P} \nu_2) \cdot \frac{2\gamma_{2,s_2}}{R^2},$$

$$\nu_1 = \frac{\nu_0^2}{\gamma_{2,s_2}^2} \ln(\frac{2^{r_{2,s_2}} + 1}{\alpha_2 - \alpha_1(2^{r_{2,s_2}} - 1)}).$$

From (18) and (19), given a first-stage coverage ratio $C_1$, the mean value and variance of minimal reception rate per unit bandwidth $R_{2,s_2}$, that supports SUs to decode their demand signals can be obtained. According to thrice standard error principle of normal distribution, the rate threshold for relay selection in large number of MG users is set to

$$R_{\sigma_0}(C_1) = F_{R_{2,s_2}}^{-1}(1 - C_1 + 3 \cdot \delta(C_1) \sqrt{N}).$$ (21)

In summary, the SINR threshold $\sigma_0$ is given for RU selection in different number of users in each MG $M_i$. When $M_i$ is small, $\sigma_0 = SINR_{S_2}$, When $M_i$ is large, $\sigma_0 = 2R_{\sigma_0}(C_1) - 1$, which is only decided by the coverage ratio in the first stage.
For $x_2$, the reception SINR of $u$ from its RU $SU_{i,m_1}$ is

$$SINR_{m_1,u} = \frac{\alpha_2 P_D |h_{m_1,u}|^2 \gamma_{m_1,u}^{-\beta}}{\alpha_1 P_D |h_{m_1,u}|^2 \gamma_{m_1,u}^{-\beta} + 2N_0}$$

$$= \frac{\alpha_2 P_D |h_{m_1,u}|^2 \gamma_{m_1,u}^{-\beta}}{\alpha_1 P_D |h_{m_1,u}|^2 \gamma_{m_1,u}^{-\beta} + 2},$$

(26)

with $x_{m_1,u} = \frac{P_D}{2N_0} |h_{m_1,u}|^2 \gamma_{m_1,u}^{-\beta}$. Let $y_{m_1,u} = \frac{\alpha_2 x_{m_1,u}}{\alpha_1}$. CDF of the instantaneous SNR $x_{m_1,u}$ is given by

$$F_{x_{m_1,u}}(x_{m_1,u}) = 1 - \exp(-\lambda_{m_1,u} x_{m_1,u}),$$

(27)

with exponential distribution parameter $\lambda_{m_1,u} = \frac{\gamma_{m_1,u}}{\rho_1}$. So the conditional CDF of $Y_{m_1,u} = \frac{\alpha_2 x_{m_1,u}}{\alpha_1} + \frac{2}{\alpha_2 - \alpha_3 y_{m_1,u}}$ is

$$F_{Y_{m_1,u}}(y_{m_1,u} | \gamma_{m_1,u}) = 1 - \exp \left(-\frac{y_{m_1,u}}{\gamma_{m_1,u}} \frac{2}{\rho_1} \right).$$

(28)

The probability that the SINR provided by the selected RU is higher than $\sigma_0$ is

$$P(SINR_{m_1,u} > \sigma_0 | \gamma_{m_1,u}) = 1 - F_{Y_{m_1,u}}(y_{m_1,u} | \gamma_{m_1,u}) \sigma_0 | \gamma_{m_1,u})$$

(29)

At the same time, $SU_{i,m}$ must be within the ETR with radius $R_0$ of RU $u$. Due to $R_0 << R$, the circle transmission region where $SU_{i,m}$ may exist can be approximated to its circumscribed sector field, as is shown in Fig.3. The probability that there is a SU being selected as $u$’s RU is

$$P_m(\gamma_u) = \theta(\gamma_u) \int_{\gamma_u - R_0}^{\gamma_u + R_0} P(SINR_{m_1,u} > \sigma_0 | \gamma_{m_1,u})$$

$$\cdot f_\gamma(\gamma_{m_1,u}) d\gamma_{m_1,u},$$

(30)

where $\theta(\gamma_u) = \arcsin \left( \frac{\gamma_u}{R} \right)$ and $\gamma_u$ is the distance between $u$ and BS. Because of the uniform distribution of MG users in the cell, the joint PDF of the distance from the $SU_{i,m_1}$ to $u$ is expressed by $f_\gamma(\gamma_{m_1,u}) = \frac{2\gamma_{m_1,u}}{R^2}$. Aiming at any one USU, it is assumed that the distance from BS is larger than $R_0$. The average probability that USU is served by one RU

$$P_m = \frac{1}{2\pi} \int_{R_0}^R P_m(\gamma_u) f_\gamma(\gamma_u) d\gamma_u.$$
From the analysis above, it is hard to find the close-form expression of $P_m$. We obtain its asymptotically tight approximation in (36), when $\rho_1 \rightarrow \infty$. Where

$$
P(SINR_{m,u} > \sigma_0|\gamma_{m,u}) = \exp\left(-\frac{(\gamma_{m,u})^\alpha}{\rho_1} \frac{2\sigma_0}{(\alpha_2 - \alpha_1\sigma_0)}\right)$$

(34)

$$
\approx 1 - \frac{(\gamma_{m,u})^\alpha}{\rho_1} \frac{2\sigma_0}{(\alpha_2 - \alpha_1\sigma_0)} \approx 1.
$$

Substituting (36) into (32), after simple algebraic operation, $C_2$ is derived. The approximate $C_{COM-NOMA}^A$ is expressed by.

$$
C_{COM-NOMA}^A \approx C_1 + (1 - C_1) \left(1 - (1 - P_m^A)^{C_1(M_1 + M_2)}\right).
$$

(35)

B. Analysis of Complexity

The complexity of the proposed COM-NOMA scheme is analyzed in the following. In the transmission process, there are three main steps affecting the complexity. The first step: the channel gains of users in the two multicast groups $MG_1$ and $MG_2$, are sorted in descending order, separately. The complexity is $O(M_1 \log_2 M_1) + O(M_2 \log_2 M_2)$. The second step: through MMGM, $\mathcal{H}_S$ is divided into three subgroups, $\mathcal{H}_{S_{ab}}$, $\mathcal{H}_{S_{m}}$, and $\mathcal{H}_{S_{u}}$. According to MMGM, it takes up to $M_1$ comparisons to divide $\mathcal{H}_S$ into three subgroups. The complexity is equal to $O(C_1 M_1)$. The third step: in RU selection, there are $C_1(M_1 + M_2)$ SUs (they are candidate RUs) and $(1 - C_1)(M_1 + M_2)$ USUs. The RU set of USU $u$ is $\mathcal{R}_u = \{SU_i, |i, m, \gamma_{m,u} = R_u, SINR_{m,u} = \sigma_0\}$. So the complexity is $O((C_1 (1 - C_1)(M_1 + M_2)^2)$. In summary, system complexity is the sum of the complexity of these three steps.

$$
O(M_1 \log_2 M_1) + O(M_2 \log_2 M_2) + O(C_1 M_1) +
O(C_1 (1 - C_1)(M_1 + M_2)^2))
$$

(37)

$$
=O(C_1 (1 - C_1)(M_1 + M_2)^2).
$$

V. PERFORMANCE EVALUATION

To illustrate the performance of the proposed COM-NOMA scheme and verify the analysis results of system coverage ratio, we provide numerical examples in this section. The simulation parameters are shown in Table I. Another important parameter of COM-NOMA is the coverage ratio of the first stage $C_1$, $C_1 \in (0.5, 0.8]$. Because lower $C_1$ than 0.5 means that each USU can not own one relay on average. There will not be enough SUs being selected as RUs of USUs in the second stage, especially when $M_1$ is small. However, when $C_1$ becomes larger and close to 1, in the first stage, $H_{2,s_2}$ becomes smaller. As the multicast rate is limited by the user with the worst channel condition. It is the more unfair to the user with good channel condition.

To simplify simulation, it is assumed that two MGs have the same number of users, $M_1 = M_2$. In this paper, when the number of users of each MG is smaller than 60, it is considered as the small number of users. BS broadcasts the SINR threshold in each time slot. However, when it is larger than 100, it is considered as the large number of users. The SINR threshold is calculated according to (24) and BS broadcasts it at the beginning of the data transmission.

Fig. 4. SINR threshold in large number of MG user (a): SINR threshold versus $C_1$, in different number of users. (b): Probabilistic graphical of SINR threshold with $M_1 = 120$, in different $C_1$.

Fig. 4. SINR threshold in large number of MG user (a): SINR threshold versus $C_1$, in different number of users. (b): Probabilistic graphical of SINR threshold with $M_1 = 120$, in different $C_1$.

TABLE I SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage Radius of Cellular</td>
<td>1000m</td>
</tr>
<tr>
<td>Efficient Transmission Range</td>
<td>100m</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>10MHz</td>
</tr>
<tr>
<td>Transmission Power of BS</td>
<td>34dBm</td>
</tr>
<tr>
<td>Transmission Power of User Terminal</td>
<td>17dBm</td>
</tr>
<tr>
<td>Noise Power Spectrum Density</td>
<td>$-174dBm/Hz$</td>
</tr>
<tr>
<td>Path Loss Coefficient</td>
<td>4</td>
</tr>
<tr>
<td>Power Allocation Factor $\alpha_1$</td>
<td>0.2</td>
</tr>
<tr>
<td>Power Allocation Factor $\alpha_2$</td>
<td>0.8</td>
</tr>
</tbody>
</table>
A. Performance Evaluation for SINR Threshold

In Fig. 4, the analytical SINR threshold, which is computed through substituting (24) into $\sigma_0 = 2R_{\sigma_0}(C_1) - 1$, and the first-stage multicast SINRs of 1000 time slots are shown, when the number of MG users is large. From Fig. 4(a), in different $M_i$, the analytical SINR threshold is not less than the maximum value of the multicast SINR in the first stage. It indicates that the analytical SINR threshold can guarantee the successful reception of the USUs who are in D2MD group. Fig. 4(b) shows the distribution curve of $\text{SINR}_{S_2}$ in different $C_1$, when $M_i = 120$. The ordinate indicates the frequency of every $\text{SINR}_{S_2}$ value. As can be seen from the figure, the distribution interval of $\text{SINR}_{S_2}$ increases with the increase of $C_1$, which can also be verified by Fig. 4(a). The probabilities that $\text{SINR}_{S_2}$ is greater than the SINR threshold $\delta_1$ are given. They are all less than $10^{-5}$, and decrease as $C_1$ increases. Thus, the analytical SINR threshold can reduce the frequency that BS computes and decreases as $\delta_1$.

The analytical SINR threshold is reasonable for USU to select RU when $M_i$ is large.

B. Comparison with Existing Multicast Schemes

Considering that NOMA is employed, we compare the proposed COM-NOMA scheme with four other multicast schemes.

COM-OFDMA uses the same cooperation scheme with COM-NOMA scheme, but it is OFDMA-based, i.e., the two signals, $x_1$ and $x_2$, are separately transmitted on two orthogonal frequency bands. CM-OFDMA is a CM scheme proposed in [4]. The above three schemes support a fixed proportion of users in the first stage with coverage ratio $C_1$. But in CM-OFDMA scheme, USU combines receive signals from all SUs. When the rate of combined signals is higher than that of the first stage, the demand signal can be decoded. In the two OFDMA-based CM schemes, each MG collaborates in the group and the signals of two MGs are respectively transmitted on the two orthogonal frequency bands. Over the entire time slot, opportunistic multicast scheme (OMS) in [23] only supports a definite proportion of users who own good channel conditions, and its system coverage ratio is $C_1$. MS-NOMA is an NOMA-based multicast scheme serving all MG users, which is proposed in [16].

1) Power Efficiency Comparison: Due to different multicast schemes employ different number of RUs to forward data in the second stage, their total power consumption is different. For fairness, power efficiency of 5 kinds of multicast schemes are plotted in Fig. 5. The total power is denoted by $P_{\text{tot}} = \frac{1}{T}P + \frac{1}{T}P'$ [24], where $P'$ is the total power of RUs in the second stage. From Fig. 5, it can be observed that COM-NOMA performs better than other multicast schemes when $M_i$ is larger than 10. It is because that receiving both demand signals increases the probability that SUs forward signals to USUs, whichever MG they belong to. The cooperation between two MGs is achieved. When MG users are uniformly distributed in the cell, the number of candidate RUs is twice as many as other CM schemes. USU has more opportunities to be relayed by users. And each RU ensures that the USUs he (she) serves receive successfully. More D2MD groups are constituted in the second stage. Because the number of USUs is also larger than those two OFDMA-based CM schemes, the distance between USU and SU is relatively smaller. The number of members in each D2MD group is also larger. Multicast capacity is influenced by the number of users in each MG. Because $P_{\text{DS}} << P$, the power consumption of BS is the main part of the total power consumption. In these three CM schemes, the power consumption of BS is the same. COM-NOMA achieves higher power efficiency than them.

In COM-OFDMA, due to the signals of two MGs are transmitted on the orthogonal frequency bands, different MGs can not cooperate with each other. The distance between users in the same MG may be larger than the radius of ETR. Therefore, intra-group cooperation also can not be performed well. BS only transmits data in the first half of the time slot, so its power efficiency is about twice that of OMS, when $M_i$ is smaller 60. Although CM-OFDMA makes all

\[
P_{m_i}^A = \frac{1}{2\pi} \int_{R_0}^R P_m(\gamma_u) \frac{2\gamma_u}{R^2} d\gamma_u
\]

\[
\approx \frac{1}{2\pi} \int_{R_0}^R \arcsin \left( \frac{R_0}{\gamma_u} \left( \int_{\gamma_u - R_0}^{\gamma_u + R_0} \frac{2\gamma_m u}{R^2} d\gamma_{m,u} \right) \right) \frac{2\gamma_u}{R^2} d\gamma_u
\]

\[
= \frac{1}{2\pi} \int_{R_0}^R \arcsin \left( \frac{R_0}{\gamma_u} \frac{4\gamma_u R_0 2\gamma_u}{R^2} \right) \frac{2\gamma_u}{R^2} d\gamma_u
\]

\[
= \frac{2R_0}{3\pi R^4} \left( R_0 R \sqrt{R^2 - R_0^2} + 2R^3 \arcsin \left( \frac{R_0}{R} \right) + R_0^3 \ln \left( 1 + \sqrt{1 - \frac{R_0^2}{R^2}} \right) \right) - \pi R_0^3 - R_0^3 \ln R_0
\]
SUs as relays to serve USUs in the second phase, severe path loss which is caused by the large distance between SU and USU impairs its performance. Therefore, their performance is lower than COM-NOMA scheme.

However, in Fig. 5(a), which simulates the power efficiency of COM-NOMA scheme with SINR threshold $\sigma_0 = SINR_{S_2}$ under small number of MG users. When $M_i$ is smaller than about 10, the power efficiency of COM-NOMA is worse than COM-OFDMA. Because user density is lower in the cellular, SU hardly locates within the ETR of USU.

Inter-group cooperation cannot be performed well. In order to guarantee efficient transmission of selected RUs, COM-NOMA scheme ignores the signal that makes reception SINR of USU smaller than threshold. This leads to less relays than CM-OFDMA to complete the transmission in the second stage. At the same time, $P_D << P$, the power of BS is the main factor of total power consumption, as $M_i$ is small. The signal of one MG is the other MG's interference, so the performance of COM-NOMA is worse in comparison with CM-OFDMA. That also produces the
Coverage Ratio

0.95
0.85
0.75
0.7
0.6
0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9

0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9

0.75
0
.8
0.85
0.9
0.95

Coverage Ratio
Approximation
Simulation
C\textsubscript{1}=0.8
C\textsubscript{1}=0.7
C\textsubscript{1}=0.6

Fig. 7. System coverage ratio versus PAF. (a): The number of MG users is small and \(M\textsubscript{i}=40\). SINR threshold \(\sigma\textsubscript{0}=\text{SINR}_{\text{S_i}}\). (b): The number of MG users is large and \(M\textsubscript{i}=120\), SINR threshold \(\sigma\textsubscript{0}=2^{R\sigma\textsubscript{0}(C\textsubscript{1})}-1\) poor performance of MS-NOMA. In Fig. 5(b), the threshold is \(\sigma\textsubscript{0}=2^{R\sigma\textsubscript{0}(C\textsubscript{1})}-1\) for large \(M\textsubscript{i}\). CM-OFDMA has high power consumption due to all SUs are selected as relays. Its power efficiency is worse than that of OMS and COM-OFDMA.

2) Coverage Ratio Comparison: The comparison of coverage performance between COM-NOMA and COM-OFDMA under different \(C\textsubscript{1}\) is shown in Fig. 6. Since CM-OFDMA makes all SUs as the relays of USUs to forward data in the second stage regardless of device's ETR. For the sake of fairness, only the COM-OFDMA that utilizes the same relay selection scheme as the COM-NOMA is shown in the figure. Fig. 6(a) and Fig. 6(b) show the coverage performance under different number of MG users. COM-NOMA's coverage performance is superior to COM-OFDMA. Especially when \(M\textsubscript{i}=200\), it increases by about 10%. This is because COM-NOMA implements inter-group cooperation between two MGs. When MG users are uniformly distributed in the cell, users of both MGs in the ETR of the USU can receive their flag messages. The probability that SUs are selected as relays increases, and the proposed relay selection scheme ensures that each USU covered by RU can successfully receive, which also increases the coverage ratio of system. It can be seen that the approximate and simulated coverage performance match, except for \(C\textsubscript{1}=0.6\) in Fig. 6(b). When \(C\textsubscript{1}=0.6\) in Fig. 6(b), approximate coverage ratio is larger than the simulated one. The difference between the approximate and simulated coverage ratio is mainly caused by the transmission in the second stage. It is because that in the calculation of the second-stage coverage ratio, the circular area actually covered by RU is approximated to its circumscribed sector field, which increases the coverage, i.e., the approximate \(P\textsubscript{m}\) increases. The approximate expression of coverage ratio (35) is an increasing function of \(P\textsubscript{m}\), and \((M\textsubscript{i}+M\textsubscript{2})\). And when \(C\textsubscript{1}=0.6\), factor \((1-C\textsubscript{1})\) is larger than \(C\textsubscript{1}=0.7\) and \(C\textsubscript{1}=0.8\). Therefore, the difference between approximate coverage ratio and the simulated one is obvious, when the number of MG users is large and \(C\textsubscript{1}=0.6\). But the error is still less than 3% of the simulated coverage ratio.

3) Complexity Comparison: The complexity of the other four multicast schemes, i.e., COM-OFDMA, CM-NOMA, MS-NOMA and OMS, is also mainly influenced by the three steps in the Sec IV-B. The complexity analysis and comparison is shown in Tab. II. From Tab. II, the complexity of COM-NOMA is higher than that of the other schemes. It is because that it performs RU selection for USU in the second stage after the first-stage opportunistic multicast with a fixed coverage ratio. But it improves the main drawback of OMS, namely, poor fairness. In OMS, USUs have no chance to recept successfully, due to their poor channel conditions, even though they have already paid for the content.

The complexity of COM-NOMA is determined by \(C\textsubscript{1}\), and the total number of users in the two MGs, \(M\textsubscript{i}+M\textsubscript{2}\). The complexity of those two OFDMA-based CM schemes, COM-OFDMA and CM-OFDMA, is determined by \(C\textsubscript{1}\) and \(M\textsubscript{i}\). It is because that they only perform collaboration within the single MG, unlike COM-NOMA, which also supports the collaboration between two MGs. As a result, in the third step, the amount of candidate RUs is less than the COM-NOMA for each MG.

Even though the complexity of the COM-NOMA is higher than other schemes, its energy efficiency is greatly improved. For example, in Figure 5(a), when \(M\textsubscript{i}=60\), the complexity of COM-NOMA is twice that of COM-OFDMA and CM-
OFDMA, but the power efficiency of COM-NOMA is 3 times that of COM-OFDMA, and 5 times that of CM-OFDMA. The complexity of COM-NOMA is 5 times that of MS-NOMA, but its power efficiency is 12 times that of MS-NOMA.

C. Impact of PAF on Coverage Ratio

As an important parameter of NOMA, the impact of different PAFs on the coverage ratio of COM-NOMA is simulated in Fig.7. Coverage ratio is nearly invariable under different PAFs. It is because that the coverage ratio \( C_1 \) is fixed (e.g., 0.6, 0.7, 0.8) in the first stage. The improvement of system coverage ratio is produced by the cooperation among MG users in the second stage. However, in the second stage, D2MD members (i.e., those USUs who are served by RUs) are distributed in the small ETRs of RUs. The small path loss caused by the small distance between RU and D2MD member is the key factor to determine that the reception SINR of signal \( x_2 \) for D2MD members is larger than \( \sigma_0 \), even though the transmission power of RU is low. This also guarantees that each selected RU can make the USUs he (she) serves receive successfully. Therefore, coverage ratio of the second stage \( C_2 \) does not change significantly with \( \alpha_2 \), which also validates the proof in (35) and (36) that coverage ratio is independent of PAF. Therefore, the coverage ratio of COM-NOMA does not change significantly with \( \alpha_2 \).

VI. CONCLUSIONS AND FUTURE WORK

In this paper, an NOMA-based CM scheme is proposed to achieve inter-group cooperation between two MGs, which is termed as COM-NOMA. The signals of two MGs are transmitted on the same licensed spectrum simultaneously in NOMA way. Thus, every SU gets the both signals. SUs in one MG have the chance to be selected as the RUs to serve the USUs in the other MG. To guarantee that the USUs served by RUs can successfully decode, SINR threshold for RU selection is given based on the extreme value theory and the statistical property of the first-stage multicast SINR, when the number of MG users is large (e.g., 100). It is different from the case that the number of MG users is small (e.g., 60). BS doesn’t have to calculate and broadcast the first-stage multicast SINR in every time slot. We have further investigated the coverage ratio of COM-NOMA. The approximated close-form expression of system coverage ratio is derived. According to the analytical result, it is mainly impacted by number the number of multicast users. The PAF, which is an important parameter of NOMA, scarcely contributes to coverage ratio. Simulation and numerical computations have been carried out to verify the performance of COM-NOMA. It shows that power efficiency enhances when the number of users is larger than 10. Moreover, the coverage ratio enhances in comparison with COM-OFDMA, especially when \( M_i = 200 \). It increases by about 10%. And the analytical coverage ratio and the simulated one match well.

In the future, the boundary of the total number of MG users will be investigated to choose proper SINR threshold, which is not investigated in this paper. In addition, the proposed COM-NOMA schemes is not energy-efficient, because one USU can select multiple SUs as his (her) RUs (especially in the area close to BS) and every RU can make sure successful reception. According to the user density in the cell and the distance between user and BS, the optimal RU is selected to save energy.

REFERENCES


